Dept. of Math. University of Oslo
Statistical Research Report No. 12
ISSN 0806–3842 December 2004

# QUANTUM THEORY AS A STATISTICAL THEORY UNDER SYMMETRY AND COMPLEMENTARITY.

Inge S. Helland
Department of Mathematics,
University of Oslo,
P.O. Box 1053 Blindern,
N-0316 Oslo, Norway
E-mail: ingeh@math.uio.no
Ph. (047)22855873

Running head: Quantum theory and statistical theory

**Abstract**

The aim of the paper is to derive essential elements of quantum mechanics from a parametric structure extending that of traditional mathematical statistics. The main extensions relate to symmetry, the choice between complementary experiments and hence complementary parametric models, and use of the fact that there is a limited experimental basis that is common to all potential experiments. Concepts related to transformation groups together with the statistical concept of sufficiency are used in the construction of the quantum mechanical Hilbert space. The Born formula is motivated through recent analysis by Deutsch and Gill, and is shown to imply the formulae of elementary quantum probability/ quantum inference theory in the simple case. Planck's constant, and the Schrödinger equation are also derived from this conceptual framework. The theory is illustrated by one and by two spin 1/2 particles; in particular, a statistical discussion of Bell's inequality is given.

Key words: Parameters (state variables); Quantum theory; Statistical models; Statistics (observations); Symmetry.
AMS Subject Classification: 81P10, 62A99

## 1 Introduction.

Nobody doubts the correctness of quantum mechanics. But the completeness of the theory has been debated since Einstein, Podolsky and Rosen [1] first raised the issue explicitly in 1935. Consider an analogy: A well known general theorem by Gödel from 1931 - see [2] - says that every rich enough theory may be regarded as incomplete in a certain sense. The difficulty with quantum theory is that it is nearly impossible to discuss simply its bordering area, say towards macroscopic theories in general or towards relativity specifically, since the foundation of the theory is always presented in purely formal terms.

Some intuitive notions have of course been developed around quantum mechanics during the years, but it is very difficult to have any immediate understanding of a theory starting by stating that an observable - whatever that should mean in intuitive terms - is defined as a selfadjoint operator on a complex Hilbert space. This is of course an important element of quantum theory, but taken as a basis for axiomatics, it is very formal.

The other foundation stone of modern physics, special relativity, has a beautiful basis of simple assumptions. Note that 'simple' here does not mainly mean simple in formal mathematical terms, but rather in everyday language: physical laws are the same for all observers, and the speed of light is the same. So the question is: Is there some possibility of finding a similar simple basis for quantum mechanics also?

The main purpose of this paper is to suggest a foundation of quantum mechanics based on relatively simple concepts like: choice of experiment, statistical parameter, symmetry and model reduction. We claim that this approach may lead to a conceptual starting point which is more intuitive than the usual one. Parts of our goal will also be to develop a theory which brings the statistical tradition and the traditions developed in physics closer, ultimately to the extent that the two traditions may learn from each other. It should be unnecessary to point out, of course, that with an aim as ambitious as that, there will be open questions, both technical ones and questions related to the underlying philosophy and to the interpretation of concepts. The hope is, however, that the process started here will continue, and that this process in the end will turn out to be of some benefit to both sciences.

It is well known that there exist a large number of interpretations of quantum theory; an incomplete list is given by the references [3, 4, 5, 6, 7]. The present article implies a particular statistical interpretation closely related to the epistemic view of states [8], to Bohr's original minimalistic view and also to the neo-Copenhagen interpretation [9]. Our main focus, however, will be on trying to derive the theory using simpler, less formal concepts. This will be done through a thorough discussion of the structure of the parameter spaces of the relevant experiments.

There are a few related papers in the recent literature. A. Bohr and Ulfbeck [10] discuss a foundation of quantum mechanics which is based upon irreducible representation of groups, and thus uses symmetry in a way which is similar to ours. Caves et al [11] proposes a Bayesian approach to quantum theory based upon Gleason's powerful Hilbert space theorem. Here we will avoid taking an abstract Hilbert space as a point of departure, but we will arrive at it from a rather concrete setting. Finally, Hardy [12] derives quantum theory and probability theory in an elegant way from a few reasonable axioms, but using the concept of measurement in a simpler way than we do here. It must be emphasized that we go further in looking at complementary models and in reducing models under symmetry than what is common in the present statistical literature. One hope is that this later can be justified more explicitly from a prediction point of view.

A completely different approach towards the same aim as the present paper, using formal quantum lattice theory, is discussed in [13].

What are then our conclusions concerning the completeness of quantum mechanics? According to our view, quantum theory itself can be interpreted as a statistical theory, and is as such reasonably complete. The corresponding model parameter may in some sense be related to a hidden variable of the kind first rejected by von Neumann [14], but later defended by Bell [15] and others. However, in our view a hidden parameter is a simpler, much more flexible and also more adequate concept. Below we will introduce the concept of *total parameter*, a pure modelling concept which may comprise several potential experiments. A total parameter will not in general take a value, in agreement with the Kochen-Specker theorem, but also in agreement with the fact that there is a limit to how many parameters you can make inference on from limited data in an ordinary statistical experiment.

A basic attitude behind the present paper is the following: Physics is an empirical science, and seeking its foundation one should look at methods and model considerations that have proved useful in other empirical sciences. In my opinion, too much of this field is dominated by formal mathematics. Mathematics is of course important and useful, but the very foundation of physics should be simple, and one should then refer to a concept of simplicity which is based on empirical science, not necessarily on notions belonging to the mathematical tradition.

The plan of the paper is as follows: The background and basic concepts are introduced in Sections 2-7. In Sections 8-10 the one- and two-particle situations are discussed, including a statistical treatment of Bell's inequality. A survey of group representation theory and related concepts are given in Sections 11-12. Then in Sections 13-15 a construction of the basic Hilbert space for quantum mechanics is made using statistical concepts and symmetry. This is probably the most important contribution of the present paper. Various arguments for the Born formula are briefly discussed in Section 16, and in Section 17 it is indicated how essential elements of quantum mechanics and of quantum statistcs may be deduced from this. Section 19 discusses the Lorentz transformation and Planck's constant, while an argument for the Schrödinger equation is given in Section 20.

## 2 Statistics and quantum theory.

As is well known, statistical methodology has had applications in most areas of empirical science, including experimental physics. Statistical inference is based upon a relatively simple paradigm: There is an unknown part of reality that we want to learn something about; this is described by a *parameter* $\theta$. Learning is done through making observations $y$, and in general the act of making such observations is called an *experiment*. A model for an experiment is made by postulating probability measures on a sample space $S$, that is, a space connected to potential observations. The model is then given by a class of probability measures on $S$; say $\{\mathrm{P}^\theta(\cdot)\}$, that is, the measures are indexed by the unknown assumed part of reality $\theta$. The observations $y$ are stochastic variables, i.e., functions on the sample space $S$.

Statistical inference is the art of deducing information on $\theta$ from the observations $y$. The Bayesian school of statistical inference theory assumes in addition to the model a prior distribution on the parameter space.

The Bayesian way of thinking also has a strong position in current quantum information theory, see Fuchs [8] and references there. If we interpret the statistical parameters as something like quantum theoretical state variables, which will be the point of departure for much of the present paper, the distinction between quantum information Bayesianism and statistical Bayesianism will be relatively small. In statistical theory there exist viewpoints labeled 'objective Bayesianism' [16], which may sound like a contradiction, but which can in fact be made to make sense. One version of this, where priors are induced by symmetry groups, in fact underlies much of the present paper.

An important distinction between quantum information and statistics is that the main application of the Bayesian assumption in statistics is in the inference from observations to parameters using the measurement model, that is, the measure $\mathrm{P}^\theta(\cdot)$ on the potential observations. This statistical measurement model is currently used routinely and with success in a large number of sciences, medicine, biology, social sciences and so on. In this paper we will emphasize the difference between observations $y$ and state variables $\theta$. There is no reason why such a statistical point of view shouldn't be relevant to physics also. Also in physics any measurement apparatus implies uncertainty. In fact we intend to show in this paper that it may be very fruitful for the understanding of quantum theory to regard state variables as statistical parameter determining the distribution of the observations. The tradition in quantum physics has been to concentrate on other aspects of the measurement process, namely those envisaged by von Neumanns formal analysis. On the other hand, ideas related to ours are common in the operational approach to quantum physics and in quantum information.

We will come back to the measurement model and the observations later in this paper; first we will concentrate on the state variables or parameters. These parameters will be important throughout the paper.

It will be a point of departure that the parameter as such only makes sense within the experiment that it is connected to: A certain question is raised by the experiment, and the value of the parameter is the ideal answer to that question. A new element will later turn up, though, as a consequence of the theory: Given the value of the parameter, then the symmetry assumptions, the assumption of model reduction and the fact that the same unit always is involved, seems to imply the well known Born formula, amounting to the following: If another experiment is done on this unit, a probability distribution of the relevant parameter for this latter experiment can also be found. In fact, it is here that the quantum formalism supplies completely new elements to ordinary statistical inference, valid under the stringent symmetry conditions that one finds in the simple microscopic world. The statistical argument behind Born's formula will only be indicated in this paper by summarizing results by Deutsch, Gill and others.

The distinction between the parameter space defining the state of a system and the sample space (space of observations) with its estimators is essential for the present paper. The distinction is not usually made in physics, but it is crucial in statistical inference. We will keep the distinction even in cases with perfect observation, where the value of each observation almost equals some function of the parameter. This is consistent with current statistical theory, and it may be a way to understand better certain paradoxes of quantum mechanics. We will also extend the parameter space (state space) to contain state variables that are known or assumed known, contrary to what is common in statistics.

An essential point of the statistical paradigm is that, before the experiment is done, the parameter $\theta$ is unknown; afterwards it is as a rule fairly accurately determined. In this way the focus is shifted from what the value of the parameter 'is' to the knowledge we have about the parameter. In a physical context this can easily be made consistent with the point of view expressed by Niels Bohr [17]: 'It is wrong to think that the task of physics is to find out how nature is. Physics concerns what we can say about nature.'

In several cases the statistical model may be too rich for the parameters to be identified by estimation, but even so the parameters may be of interest. On certain occasions there may be a choice with respect to which parameter to estimate. For instance, assume that we want measure some quantity with an apparatus which is so fragile that it is destroyed after a single measurement. We may model the measured values to have an expectation $\mu$ and a standard deviation $\sigma$, perhaps even a normal distribution with these parameters. A single measurement gives an estimate of $\mu$. The standard deviation may be thought to be possible to estimate by dismantling the apparatus, again destroying it. This then gives a first example with two complementary parameters: Only one of them can be measured.

Important ingredients of the paper seen from a statistical point of view are: Model reduction, symmetry, complementary parameters.

The last of these concepts is particularly important, and extends the common statistical way of thinking. In a model of a particle we can imagine that it has a theoretical, definite position $\theta^1 = \xi$ and a theoretical momentum $\theta^2 = \pi$, but there is a limit to how accurate these parameters can be determined. From our point of view, this is conceptually not much more difficult than the following:

A given patient has (expected) recovery time $\theta^1$ if treatment 1 is used and $\theta^2$ if treatment 2 is used. The term expected here must be interpreted in some loose sense, not necessarily with respect to a well-defined probability model. Like all parameters, $\theta^1$ and $\theta^2$ can be estimated from experiments, but it is impossible to estimate both parameters on the same patient at the same time. In statistics this and similar problems are solved by investigating several units, here patients, assuming the same parameters for all units. Such an assumption is relevant to ordinary statistical investigations, where the purpose is to say something of importance for a large population of patients, and hence to future patients. In quantum physics the parameter must be connected to a single particle or to a small number of particles, and then the analogy with a model for a single patient becomes of interest. As will turn out, the only possibility then of being able to infer something about such a parameter, is to make stringent symmetry assumptions.

# 3 Statistical models and groups.

Define the parameter $\theta$ of an experiment as above, and let a symmetry group $\bar{G}$ be defined on the parameter space. This group $\bar{G}$ will be kept fixed, being thought of as a part of the specification of the model. The basic requirement for choosing $\bar{G}$ is that the parameter space should be closed under the actions $\bar{g}$ of the group: $\theta \mapsto \theta\bar{g}$, where it is convenient to place the symbol for the group element acting on $\theta$ on the right. (This will lead naturally to the right-invariant measure as the non-informative prior on the parameter space, a solution that was argued for in [19] to be the best one from several points of view.)

Throughout this paper, we will regard groups as *transformation groups* acting on concrete spaces, primarily the parameter space, but also the space of observations. In the mathematical literature this is called a group action, which can be regarded as a group of automorphisms of a given space.

Sometimes in statistics a symmetry group $\dot{G}$ on the sample space is defined first, and then $\bar{G}$ is introduced via the statistical model by defining $\mathrm{P}^{\theta\bar{g}}$ by

$$\mathrm{P}^{\theta\bar{g}}(A) = \mathrm{P}^{\theta}(A\dot{g}^{-1}) \ \mathit{for\ sets}\ A. \tag{1}$$

Then the connection from $\dot{G}$ to $\bar{G}$ is a *homomorphism*:

$$\dot{g}_1, \dot{g}_2 \mapsto \bar{g}_1, \bar{g}_2 \ \ \mathit{implies}\ \ \dot{g}_1\dot{g}_2 \mapsto \bar{g}_1\bar{g}_2. \tag{2}$$

The concept of homomorphism will be fundamental to this paper. It means that we have very similar group actions: The identity element, inverses and subgroups are mapped as they should from $\dot{G}$ to $\bar{G}$, i.e., the essential structure is inherited. If $\dot{g} \mapsto \bar{e}$ implies $\dot{g} = \dot{e}$ (identity elements), the homomorphism will be an *isomorphism*: The structures of the two groups are then essentially identical. If in addition a one-to-one correspondence can be established between the spaces upon which the groups act, everything will be equivalent.

# 4 Total parameters.

The model description above demands in principle that the parameter should be estimable from the available data. For models involving only one or a few units, this is typically not realistic at all, as already noted.

A simple example is the location and scale parameters $\mu, \sigma$ in a case where only measurement on a single unit is possible. As another example, assume that two questions are to be asked to an individual, and we know that the answer will depend on the order in which the questions are posed. Let $(\theta^1, \theta^2)$ be the expected answer when the questions are posed in one order, and $(\theta^3, \theta^4)$ when the questions are posed in the other order. Then $\phi = (\theta^1, \theta^2, \theta^3, \theta^4)$ cannot be estimated from one individual. Many more realistic, moderately complicated, examples exist, like the effect of treatments on a patient where only one treatment can be given, or behaviourial parameters of a rat taken together with parameters of the brain structure which can only be measured if the rat is killed.

In all these cases the situation can be amended through investigating several individuals, but this assumes that the parameters are identical for the different individuals, a simplification in many cases. Note that the ordinary statistical paradigm in the simplest case assumes an infinite population of units with the same parameters. This will not be assumed in the present paper.

When considering these cases where $\phi$ cannot be estimated from any experiment on the given units, we may call $\Phi = \{\phi\}$ a total parameter space rather than a parameter space. We nevertheless insist that modelling through total parameters can be enlightening, in the cases mentioned above as in other cases. In particular this may be useful if one models cases where one has the choice between several measurements, as one usually will have in quantum mechanics. As will be discussed below, by choosing a particular experiment in a given setting, what one can hope for, is to be able to estimate a part of the total parameter.

Sometimes it will be convenient to use the term 'parameter' both for total parameters and ordinary parameters, and then use the specific term 'estimable parameter' for the latter. The term estimability, as used here, is the same as in statistics [18]:

**Definition 1.**
*A parameter $\theta$ is unbiasedly estimable if there exists an experiment with model $\mathrm{P}^\theta(\cdot)$ and a random variable $y$ of that experiment such that*

$$\mathrm{E}^\theta(y) \equiv \int y(\omega)\mathrm{P}^\theta(d\omega) = \theta.$$

*In that case we say that the parameter $\theta$ can be estimated unbiasedly by the statistic $y$. More generally, a parameter $\theta$ is estimable if there is a one-to-one function of $\theta$ which is unbiasedly estimable.*

The last generalization is made to ensure that a one-to-one transformation of a parameter $\xi(\theta)$ is estimable whenever $\theta$ is estimable.

The total parameter space $\Phi$ can in general have almost any structure; we will assume here that it is a locally compact topological space. We will also assume that there is a transformation group $G$ acting on $\Phi$, and that $G$ satisfies certain weak technical requirements (see [19]) so that $\Phi$ can be given a *right-invariant measure* $\nu$, satisfying $\nu((d\phi)g) = \nu(d\phi)$. The invariant measure is unique for *transitive* groups, i.e., groups having the property that for each $\phi_1, \phi_2$ there exists a $g$ such that $\phi_1 g = \phi_2$. In general the invariant measure is unique on *orbits*, i.e. sets of the form $\{\phi g : g \in G\}$. It must be supplemented by a measure on the orbit indices in order to give a measure on the whole space $\Phi$.

# 5 Choice of question.

We will propose here a general procedure in physics quite similar to that often used in applied statistics: After the situation has been clarified in terms of a parametric structure, the first issue is to choose what we are interested in, and then which experiment to perform. This will then first lead to a focus parameter $\theta^a$.

There are usually many questions that can be investigated in a given setting. Typically the different such questions are addressed performing different experiments on the specific part of reality in question.

Let $\mathcal{A}$ be the set of such questions. This gives for each $a \in \mathcal{A}$ a focus parameter $\theta^a = \theta^a(\phi)$ (possibly a vector). Depending upon the circumstances, this may still be a total parameter. To achieve a crisp probability model, and through that hopefully an estimable parameter, some model reduction may have to be performed; see Section 7.

When a group $G$ is defined on the original total parameter space $\Phi$, an important property of the focus parameter $\theta^a$ is if it is a <u>permissible</u> function $\theta^a(\phi)$, that is, satisfying:

$$\text{If } \theta(\phi_1) = \theta(\phi_2) \text{ then } \theta(\phi_1 g) = \theta(\phi_2 g) \text{ for all } g \in G.$$

The most important argument for this restriction is that it leads to a uniquely defined group $\bar{G}$ on the image space $\Theta$ of $\theta(\phi)$:

$$(\theta\bar{g})(\phi) = \theta(\phi g). \tag{3}$$

Some additional arguments for the requirement of permissibility are given in [19, 21]. Among other things certain paradoxical conclusions related to Bayes estimation are avoided if focus parameters are required to be permissible functions. Thus what we do here, is to demand nature to avoid certain paradoxes.

Trivially, the full parameter $\theta = \phi$ is permissible. Also, the vector parameter $(\theta_1, \ldots, \theta_k)$ is permissible if each $\theta_i$ is permissible. These two facts indicate that in addition to the requirement that the experimental parameter $\theta^a$ should be permissible, one must typically also require that it is not too big. If necessary for estimability, model reduction must be done, as discussed below. This depends upon the context of the experiment.

As a simple illustration of a group connected to a parameter space or the total parameter space, look at the total parameter $\phi = (\mu, \sigma)$ with the translation/scale group $(\mu, \sigma) \mapsto (a + b\mu, b\sigma)$ where $b > 0$. The following one-dimensional parameters are permissible: $\mu$, $\sigma$, $\mu^3$, $\mu + \sigma$, $\mu + 3\sigma$, and if a focus parameter is asked for, all these give valid candidates.

On the other hand, the following parameters are not permissible, and would according to McCullagh [20] lead to absurd focus parameters under this group: $\mu + \sigma^2$, $\sigma e^\mu$, $\mathrm{tg}(\mu)/\sin(\sigma)$.

A further example is given by the coefficient of variation $\sigma/\mu$. This is not permissible since the location part of the transformation does not make sense here. But it will be permissible if the group is reduced to the pure scale group $(\mu, \sigma) \mapsto (b\mu, b\sigma)$, $b > 0$. Going back to the 'absurd' examples above, we also see that the first two of them will be permissible if the group is reduced to the pure translation group $(\mu, \sigma) \mapsto (a + \mu, \sigma)$. This points at an important general principle:

*If a focus parameter $\theta^a(\phi)$ is not permissible with respect to the basic group $G$, then take a sub-group $G^a$ so that it becomes permissible with respect to this subgroup.*

One can easily show in general [22] that there exists a maximal subgroup $G^a$ having the property that $\theta^a(\phi)$ is permissible with respect to this group. Then this induces a group $\bar{G}^a$ on $\Theta^a = \theta^a(\Phi)$, and there is a simple homomorphism from $G^a$ to $\bar{G}^a$ determined as in (3).

A simpler situation where the theory of this paper also applies in principle, is when the parameter set $\{\theta^a; a \in \mathcal{A}\}$ is given at the outset, together with groups $\bar{G}^a$. Then one may just define $\phi = (\theta^a; a \in \mathcal{A})$, that is, the vector of all $\theta^a$ and define $G$ by $\phi g = (\theta^a \bar{g}^a; a \in \mathcal{A})$.

# 6 Context.

Any experiment is done in a context, that is, for some given experimental units, some preconditioning done on these units, some assumptions explicitly made and verified before the experiment and some physical environment chosen for the whole experiment.

A context may include the knowledge of parameters estimated with certainty in earlier experiments. (The concept of *ideal experiment* also turns out to be useful in the last part of this paper.)

As discussed later, parts of the context may be formed by conditioning upon random variables with distribution independent of any parameters, in particular upon the experiment chosen.

In general, the context may consist of a complex of conditions upon which all probabilities for potential experiments depend.

As already discussed in the Section 2 and 4, for a given context, certain pairs of experiments are *incompatible*: Only one of these experiments can be performed in a given context. Niels Bohr used the concept of *complementarity* in a sense closely related to this. Many physicists have followed Bohr's use of the word complementarity even though this is somewhat problematic: The same word is used in a different meaning in psychology and in color theory. Some mathematical physicists, among them L. Accardi, argue that the word complementarity should only be used for potential experiments that are maximally incompatible in some precisely defined sense: For two discrete parameters, in a state determined by fixing one of them, the posterior probability distribution of the other should be uniform over its values. By taking limits, a similar notion can be defined for continuous parameters, even parameters like expected position or momentum, taking values on the whole line.

The rôle of the context in this paper will be to limit the set of random variables (statistics) that can be connected to each of several incompatible experiments. Specifically, for each experiment $a$, see below, the Hilbert space formed by functions of the complete sufficient statistic (Definition 3) for the parameter of this experiment is assumed to be non-trivial, and these Hilbert spaces are tied together by a symmetry assumption.

# 7 Model reduction.

Nearly every useful model of reality is a simplification. A simplification may sometimes be necessary if it shall be possible to perform a meaningful experiment in a given context.

**Definition 2.**

*We say that the parameter $\theta$ has been reduced to the parameter $\lambda$ if we first have $\lambda = \eta(\theta)$ and possibly also reduce the range $\Theta$ of $\theta$, and then let the statistical model be reduced to only depend upon $\lambda$. Equivalent: The reduced parameter subspace $\Lambda$ is a proper subset of $\Theta$, and the reduced parameter $\lambda$ runs through $\Lambda$.*

By suitably defining $\lambda$, this includes many cases, like equating parameter values for different individuals, letting $\lambda$ be a discretized version of $\theta$ or letting $\lambda$ be by a selected set of orbits of $\theta$ under the group $\bar{G}$.

We do not have any ambition here of formulating a complete theory of model reduction.

In this paper every model reduction corresponding to a focus parameter $\theta^a$ is done by reducing the numbers of orbits of $\bar{G}^a$ as acting on $\Theta^a$. In the extreme case the parameter space is reduced to a single orbit. There is in fact two strong argument for this policy:

1) All models should have a parameter space which is invariant with respect to the relevant group, so also for the reduced model. The only way to achieve this, is to let $\Lambda^a$ be a set of orbits of the relevant group $\bar{G}^a$.

2) Within orbits, an optimal parameter estimator exists in what is called the Pitman estimator, which is the Bayes estimator with the invariant measure as prior. Hence the only room for useful model reduction, that is, model reduction leading to better predictions, is in the orbit index.

A model reduction from $\theta^a$ to $\lambda^a$ via the orbit index, will always be permissible, which is easy to see. Also, the same group $\bar{G}^a$ can be used on the image of $\lambda^a$ as was used on the image of $\theta^a$. After having reduced the model parameter from $\theta^a$ to $\lambda^a$, we now assume that there is a measurement model

$$\mathrm{P}^{\lambda^a}(\cdot)$$

for the potential observations.

Which model reduction that is chosen, will also depend on the experimental basis, i.e., on parts of the context. We will discuss this later in a group representation setting.

A relevant point here is that essentially the same kind of model reduction has turned out to be of interest in a completely different setting: One main purpose of model reduction in statistics is to improve the predictive power of a model with too many parameters. In particular, the strategy of reducing a model through the orbit index of a suitable group, has proved to be very useful in the field of chemometry. Specifically, such a model reduction in a random regression model with the rotation group on the parameter space of regression coefficients and $x$-covariance matrix was discussed in [23]. This turned out to give relations to certain chemometric regression methods which have proved to be useful, and which have originally been motivated in more intuitive ways.

# 8 One particle model.

The statistical modelling concepts introduced so far, are rather straightforward, but they do have implications, as the following example shows. Consider a particle with a theoretical spin $\phi$ given as some vector with norm $\gamma$, and let the group $G$ be the group of rotations of this vector. Assume a basic contextual setting such that the most we can hope to be able to measure is the angular momentum component $\theta^a(\phi) = \gamma\cos(\alpha)$ in some direction given by a unit vector $a$, where $\alpha$ is the angle between $\phi$ and $a$. Here $a$ can be chosen freely. Given $a$, and given the measurement in the direction $a$, the rest of the total parameter $\phi$ will be totally unavailable.

The function $\theta^a(\cdot)$ is seen to be non-permissible for fixed $a$: Two vectors with the same component along $a$ will in general have different such components after a rotation. The maximal group $G^a$ with respect to which $\theta^a(\cdot)$ is permissible, is the group of rotations of a given vector around the axis $a$ possibly together with a $180^{\circ}$ rotation around any axis perpendicular to $a$.

The group motion induced by $G^a$ on the image space for $\theta^a$ is called $\bar{G}^a$. This group has several orbits: For each $\kappa \in (0, \gamma]$, one orbit is given by the two values $\theta^a = \kappa$ and $\theta^a = -\kappa$. In addition there is an orbit for $\kappa = 0$.

We want in general that any reduction of the parameter space should be to an orbit or to a set of orbits. This gives the maximal possible reduction $\lambda^a$ of $\theta^a$ to a single orbit $\{-\kappa, \kappa\}$. The value of $\kappa$ is basically arbitrary; we take $\kappa = 1$ to conform to the usual notation for spin 1/2 particles. The group $\bar{G}^a$ has just two elements, and can trivially be represented on a two-dimensional space. Anticipating concepts that will be introduced later, this indicates that there is a corresponding state space $\mathcal{H}$ which is two-dimensional and invariant with respect to this $G^a$. Since this should then hold for all $G^a$, it follows that $\mathcal{H}$ should be invariant with respect to the whole rotation group; thus we are lead to the usual Hilbert space for spin 1/2 particles, a two-dimensional irreducible invariant space under the rotation group.

The corresponding Hilbert spaces for particles with higher spin quantum numbers are taken in the usual way as irreducible representation spaces of the rotation group. Note that these will require several orbits of the group $\bar{G}^a$. As is well known, these representations can be indexed by the norm of the reduced spin vector.

# 9 The EPR situation and Bell's inequality.

Consider next the situation of Einstein, Podolsky and Rosen [1] as modified by Bohm, where two particles previously have been together in a spin 0 state, so that they - in our notation - later have opposite spin vectors $\phi$ and $-\phi$. In ordinary quantum mechanics this is described as an entangled state, that is, a state for two systems which is not a direct product of the component vectors. According to our programme, we will stick to the parametric description, however.

As pointed out by Bell [24] and others, correlation between distant measurements may in principle be attributed to common history, but apparently not so in this case, where Bell's inequality may be violated.

Assume that spin components $\lambda^a$ and $\mu^b$ are measured in the directions given by the unit vectors $a$ and $b$ on the two particles at distant sites $A$ and $B$, where the measured values $\hat{\lambda^a}$ and $\hat{\mu^b}$ each take values $\pm 1$. Let this be repeated 4 times: Two settings $a, a'$ at site $A$ are combined with two settings $b, b'$ at site $B$. The CHSH version of Bell's inequality then reads:

$$\mathrm{E}(\hat{\lambda}^a \hat{\mu}^b) \leq \mathrm{E}(\hat{\lambda}^a \hat{\mu}^{b'}) + \mathrm{E}(\hat{\lambda}^{a'} \hat{\mu}^b) + \mathrm{E}(\hat{\lambda}^{a'} \hat{\mu}^{b'}) + 2. \tag{4}$$

In fact we can easily show the seemingly stronger statement:

$$\hat{\lambda}^a \hat{\mu}^{b'} + \hat{\lambda}^{a'} \hat{\mu}^b + \hat{\lambda}^{a'} \hat{\mu}^{b'} - \hat{\lambda}^a \hat{\mu}^b = \pm 2 \tag{5}$$

whenever all estimates take the values $\pm 1$: All the products take values $\pm 1$ and $\hat{\lambda}^a \hat{\mu}^b$ is the same as the product of the first three similar terms. Listing the possibilities of signs here, then shows that the lefthand side of (5) always equals $\pm 2$.

As is well known, the inequality (4) can be violated in the quantum mechanical case, and this is also well documented experimentally. There is a large literature on Bell's inequality. In recent years there has been a discussion [25, 26, 27] on whether or not it is possible to break the inequality by a computer experiment. Various possible positions that may be held on the violation of the inequality are discussed in [28]. One such position is that there always will be a loophole in real experiments [29] such that the experimental violation still can be explaned by a local realistic model.

The following is an important part of our philosophy: Quantum theory is a statistical theory, and should be interpreted as such. In that sense the comparison to a classical mechanical world picture, and the term 'local realism' enherited from this comparison is not necessarily of interest. We are more interested in the comparison of ordinary statistical theory and quantum theory.

Our aim is that it in principle should be possible to describe both by essentially similar ways of modelling and inference. Thus it is crucial for us is to comment on the transition from (5) to (4) from this point of view.

As pointed out by [27], for any way that the experiment is modelled by replacing the spin measurements by random variables, there is no doubt that this transition is valid, and the inequality (4) must necessarily hold. The reason is simple: The expectation operator E is the same everywhere.

Now take a general statistical inference point of view on any situation that might lead to statements like (5) and (4). Then one must be prepared to take into account the fact that there is really 4 different experiments involved in these (in)equalities. The $\hat{\lambda}$'s and $\hat{\mu}$'s are random variables, but they are also connected to statistical inference in these experiments. What we know at the outset in the EPR situation is only that some total parameter $\pm\phi$ (possibly together with other total parameter-components) is involved in each experiment. Going from this to the observations, there are really three steps involved at each node: The components $\theta(\phi)$ are selected, there is a model reduction $\lambda = \eta(\theta)$, and finally an observation $\hat{\lambda}$. Briefly: A model is picked, and there is an estimation within that model.

# 10 Statistical models in connection to Bell's inequality.

Turn to general statistical theory: According to what is called the conditionality principle [30], a principle on which there seems to be a fair amount of concensus among statisticians, inference in each experiment should always be conditional upon the experiment actually performed.

A motivating example for this is the following, due to Cox [31]: Let one have the choice between two measurements related to a parameter $\theta$, one having probability density $f_1(y, \theta)$, and the other having probability density $f_2(y, \theta)$. Assume that this choice is done by throwing a coin. Then the joint distribution of the coin result $z$ and the measurement $y$ is given by

$$c(z)f_1(y, \theta) + (1 - c(z))f_2(y, \theta),$$

where $c(z) = 1$ if model 1 is chosen, otherwise $c(z) = 0$. Should this joint distribution be used for inference? No, says Cox and common sense: All inference should be conditional upon $z$.

In particular then, the conditionality principle should apply to the distribution of point estimators. Taking this into account, it may be argued that at least under some circumstances also in the microscopic case, different expectations should be used in a complicated enough situation corresponding to (4), and then the transition from (5) to (4) is not necessarily valid.

This is dependent upon one crucial point, as seen from the conditionality principle as formulated above: When one has the choice between two experiments, the same parameter should be used in both. How can one satisfy this requirement, say, in the choice between a measurement at $a$ or at $a'$? As formulated above, the relevant parameters are $\lambda^a$ and $\lambda^{a'}$ for the two experiments under choice.

Here is one way to give a solution: Focus on the Stern-Gerlach apparatus which measures the spin. Make a fixed convention on how the measurement apparatus is moved from one location to the other. Then define a new parameter $\lambda$ which is -1 at one end of the apparatus and +1 at the other end. By using $\lambda$ as a common parameter for both experiments under choice, the conditionality principle can be applied, and (4) does not follow from (5).

As I see it, this argument can be related to what is called the chameleon effect in several papers [25, 26, 32] by Accardi. From this point of view the effect may look rather simple and uncontroversial, but note that it here is coupled to a rather deep general principle of statistics. A more detailed discussion of this and of the related loophole theme [28, 29, 33] is beyond the scope of the present paper.

My crucial point is that the violation of the Bell inequality is not by necessity a phenomenon that makes the quantum world completely different from the rest of the world as we know it. Regarding the term 'local realistic', I don't mean to imply that any macroscopic phenomenon are nonlocal if this term is relevant. But if 'realistic' means that a phenomenon always can be described by one single model, this may be a too strong requirement.

A more explicit argument for the correlation between spin measurements, using the prior at $A$ connected to model reduction there, may be given as follows: At the outset the total parameter $\phi$ is sent to $A$ and $-\phi$ to $B$. This may be interpreted to mean that much common information is shared between the two places. The vector $\phi$ is capable of providing an answer to any question $a \in \mathcal{A}$: Is the spin in direction $a$ equal to +1 or to -1?

The observer at $A$ will have a prior on $\phi$ given by a probability 1/2 on $\lambda^a = +1$ and a probability 1/2 on $\lambda^a = -1$, where $\theta^a$ is the cosinus of the angle between $a$ and $\phi$, and $\lambda^a$ the corresponding reduced parameter taking values $\pm 1$. This is equivalent to some prior on the vector $\phi$ which has probability 1/2 of being $a + \epsilon$ and 1/2 of being $-a + \epsilon$, where $a$ is a unit vector, and $\epsilon$ is some random vector perpendicular to $a$ which is independent of $\lambda^a$ and has a uniformly distributed direction. Note that this reasonable prior on $\phi$ is found by just making the decision to do a measurement in the direction $a$ at $A$.

Now let one decide to make a measurement in the direction $b$ at the site $B$. Let $b^\perp$ be a unit vector in the plane determined by $a$ and $b$, perpendicular to $b$. Then, taking the prior at $A$ as just mentioned, $\phi$ will be concentrated on $a + \epsilon = b\cos(u) + b^\perp \sin(u) + \epsilon$ and $-a + \epsilon$, where $u$ is the angle between $a$ and $b$. Hence the component of this prior for $-\phi$ along $b$ will be $-\lambda^a \cos(u) - \epsilon \cdot b$, where the first term takes two opposite values $\pm\cos(u)$ with equal probability. The expectation of this prior component will be 0, more specifically, the component will have a symmetric distribution around 0.

Conditionally, given $\lambda^a$, this prior component will have an unsymmetric distribution, and there is a uniquely distributed parameter $\mu^b$ taking values $\pm 1$ such that $\mathrm{E}(\mu^b|\lambda^a) = -\lambda^a \cos(u)$. So, using parameter reduction to $\pm 1$ at $B$, this is the distribution obtained from the model assuming a measurement in direction $a$ at $A$. There is no action at a distance here; all information is in principle contained in the total parameter $\phi$.

Turning now to estimation, in general an unbiased estimator in statistical theory is a statistic, i.e., random variable whose expectation equals the parameter in question. Let now $\hat{\lambda}^a$ and $\hat{\mu}^b$ be unbiased estimators of $\lambda^a$ and $\mu^b$, respectively, so that $\mathrm{E}(\hat{\lambda}^a|\lambda^a) = \lambda^a$ and $\mathrm{E}(\hat{\mu}^b|\mu^b) = \mu^b$. Later we shall show the existence under reasonable assumptions of such estimators taking the correct values $\pm 1$. Then

$$\begin{aligned} \mathrm{E}(\hat{\lambda}^a\hat{\mu}^b) &= \mathrm{E}(\mathrm{E}(\hat{\lambda}^a\hat{\mu}^b|\phi)) = \mathrm{E}(\mathrm{E}(\hat{\lambda}^a|\lambda^a)\mathrm{E}(\mathrm{E}(\hat{\mu}^b|\mu^b)|\lambda^a))) \\ &= \mathrm{E}(\lambda^a(-\lambda^a\cos(u))) = -\cos(u). \end{aligned} \tag{6}$$

This correlation also determines the joint distribution of the two random variables $\hat{\lambda}^a$ and $\hat{\mu}^b$.

The discussion above was partly heuristic, but it leads to the correct answer, and it seems to be a way to interpret the information contained in the total parameter $\phi$.

It is also important that the above discussion was in terms of a reasonable parametric model. Parameters are distinctly different from random variables, in particular from random variables located in time and space. Much of our daily life imply the use of mental models, and also some form of model simplification. Quantum theory can in some sense be said to have analogies also to this world, perhaps more than to the world of classical mechanics.

The limitation of the way of thinking demonstrated in this section is twofold: First, the basic group need not be the rotation group in general. Secondly, it may not be strightforward to generalize the reasoning to the case with more than two eigenvectors. Hence we will start to build up the apparatus which we feel is necessary to treat more general cases. Ultimately, it will lead to the ordinary formalism of quantum theory.

# 11 Group representation and invariant spaces.

We assume the basic elements of group representation theory to be known; for simple treatments with physical applications see [34, 35], a mathematical treatment of finite groups is given in [36] and more advanced discussions are found in [37, 38]. As is well known, group representation is a very useful tool in applications of quantum mechanics. Here, the formal apparatus of quantum mechanics will be partly derived by considering these representations. A group representation is a homomorphism of a group onto the transformations on some vector space. In simple cases one may think of the latter as a group of matrices under multiplication. It is assumed that the vector space will be invariant under these transformations. Note then that much of the statements connected to group representation will have an analogy in the group itself, looked upon as simply a transformation on a (parameter) space.

Specifically, the regular representation $U(G)$ on $L^2(\Phi, \nu)$, where $\nu$ is a right-invariant measure for the basic group $G$ is given by

$$U(g)f(\phi) = f(\phi g). \tag{7}$$

Explicitly this implies that $U(G)$ is a group of unitary linear operators acting on $\mathrm{L}^2(\Phi, \nu)$. The group property of $U(G)$ is well known and easily verified. The same formula (7) is valid for any subspace $V$ of $\mathrm{L}^2(\Phi, \nu)$ which is invariant under the group of operators $U(G)$, i.e., such that $U(g)f \in V$ when $f \in V$ and $g \in G$.

Also, there is a natural homomorphism from $G$ to $U(G)$ given by $g \mapsto U(g)$:

$$U(g_1)U(g_2)f(\phi) = U(g_1)f(\phi g_2) = f(\phi g_1 g_2) = U(g_1 g_2)f(\phi). \tag{8}$$

This means that $G$ and $U(G)$ have similar structures, which is the first basic fact that leads from a general group to the formalism of linear operators so familiar in quantum mechanics. All calculations in quantum mechanics are currently done on the operator side. As is just indicated, looking at the parameter space and the group actions defined there can sometimes lead to a more direct understanding of the same phenomena.

For some fixed $a \in \mathcal{A}$ let now $\theta^a(\cdot)$ be a subparameter (defined on the parameter space or total parameter space $\Phi$) which is permissible with respect to a subgroup $G^a$. Let

$$V^a = \{f \in L^2(\Phi, \nu) : f(\phi) = \bar{f}(\theta^a(\phi)) \text{ for some } \bar{f}\}. \tag{9}$$

This is obviously a closed subspace of $L^2(\Phi, \nu)$. Furthermore (by the property of permissibility) it is invariant under the group of operators $U(G^a)$.

Alternatively, everything can be reduced to functions of $\theta^a$: Look at the space $\bar{V}^a = L^2(\Theta^a, \bar{\nu}^a)$ with the operators $\bar{U}(\bar{G}^a)$, where $\bar{g} \in \bar{G}^a$ is defined by $(\theta^a \bar{g})(\phi) = \theta^a(\phi g)$, where $\bar{\nu}^a$ is the invariant measure on $\Theta^a$ induced by $\nu$ on $\Phi$, and $\bar{U}$ operates on functions $\bar{f}(\theta)$ by $\bar{U}(\bar{g})\bar{f}(\bar{\theta}) = \bar{f}(\bar{\theta}\bar{g})$. This means that we have a sequence of homomorphisms/isomorphisms

$$G^a \mapsto \bar{G}^a \mapsto U(G^a)(on V^a) \leftrightarrow \bar{U}(\bar{G}^a). \tag{10}$$

Sometimes a parameter $\theta^b$ (permissible with respect to $G^b$) will be a function of a permissible parameter $\theta^a$ (permissible with respect to $G^a$). This fact is equivalent to the fact that the corresponding invariant spaces satisfy $V^b \subseteq V^a$. In particular the space $V^a$ will be the same under any one-to-one reparametrization. Also in particular, if $\lambda^a$ is given by a permissible model reduction then

$$V_\lambda^a = \{f : f(\phi) = \bar{f}(\lambda^a(\phi))\} \subset V^a \tag{11}$$

.

We will call $V^a$ and $V_\lambda^a$ *parametric invariant* subspaces of $L^2(\Phi, \nu)$. Restricting group representation to these invariant spaces correspond to first going from the total parameter $\phi$ to the relevant subparameter $\theta^a$, and then to the reduced parameter $\lambda^a$. This gives a relationship between vector spaces on one hand and parameters on the other hand. There is a similar relation between group representations (acting on vector spaces) and group actions (on parameter spaces).

# 12 Experiment, model reduction and group representation.

Let now the experimentalist have the choice between different experiments $a \in \mathcal{A}$ on the same unit(s), where the experiment $a$ consists of measuring some $y^a$, with $y^a = y^a(\omega)$ being a function on some common sample space $S$, and where the measurement process at the outset is modelled with a parameter $\theta^a$. This parameter is a part of the model-description of the units, and all the model parameters may be seen as functions $\theta^a(\phi)$ of a (meta)parameter $\phi$. It must be emphasized that the total parameter here is only a modelling concept. In ordinary statistical theory one usually imagines a situation where the model applies to a number $n$ of identical units, and one then is free to let $n$ tend to infinity. Then it is obvious that every parameter must be imagined to 'have a value'. Concretely, this means that the parameter is estimable according to Definition 1. In the quantum mechanical situation we have one or a few units, and the total parameter is explicitly connected to these units. For the latter situation it does not then necessarily make sense to let every theoretical total parameter 'have a value'. This is of course consistent with the Kochen-Specker theorem. But note that $\phi$ plays a crucial rôle in the conceptual description of the situation.

We use a common sample space $S$ for all experiments $a$, since this space can be imagined in terms of a common measurement apparatus (or apparata). For convenience, we will fix one probability measure P on the sample space $S$. Each model induces a new set of probability measures $\mathrm{P}^{\theta^a}$. These probability models for the observation may depend on the way the experiment is performed. But the parameters $\theta^a$ (and $\lambda^a$) are assumed to be the same regardless of the way the experiment $a \in \mathcal{A}$ is performed.

In the examples above, we had a situation where the experimental parameters $\theta^a(\cdot)$ were non-permissible with respect to the original group $G$. As argued above, the non-permissibility means that the symmetry group $G$ on the parameter space - for the purpose of this particular experiment - must be replaced by a subgroup $G^a$, typically different for different $a$. One can show that there always for each $a$ exists a maximal such subgroup. Since this is a proper sub-group, $G^a$ cannot be transitive on the $\phi$-space, nor then the derived group $\bar{G}^a$ on $\theta^a$. This then gives us the possibility of a parameter reduction - if this is needed - which is done by selecting one orbit or some set of orbits of this group. Such a parameter reduction will always be permissible (with respect to $\bar{G}^a$, and then also with respect to $G^a$). In general, let $\lambda^a(\phi)$ be the reduced parameter. Since the model reduction is done by orbit selection, the same group symbol $\bar{G}^a$ can be used for the group acting on its range $\Lambda^a$.

We will shortly consider group representation spaces of the group $\bar{G}^a$ acting on $\theta^a$. The following argument shows that model reduction through orbit selection gives a simple transition from $V^a$ to $V_\lambda^a$, i.e., from the parameter $\theta^a$ to the reduced parameter $\lambda^a$. This gives a new model $\mathrm{P}^{\lambda^a}$, constructed from the original model $\mathrm{P}^{\theta^a}$.

Every function of a parameter $\theta^a$ can be written as a sum of functions of the $\theta^a$-parts restricted to the orbits as follows. For orbit $\mathcal{O}_i^a$ define $f_i$ by $f_i(\theta^a) = f(\theta^a)$ when $\theta^a \in \mathcal{O}_i^a$, otherwise $f_i(\theta^a) = 0$. Then

$$f(\theta^a) = \sum_i f_i(\theta^a).$$

But the set of functions of $\theta^a$-parts belonging to orbits is invariant under the relevant group $\bar{G}^a$, hence this implies a splitting into invariant spaces.

From this, the sum of subrepresentations in question corresponds to a selected union of orbits, which again corresponds to a selected reduced parameter $\lambda^a$. The statistical model with this reduced parameter will now be fixed.

# 13 Experimental basis and the Hilbert space of a single experiment.

Up to now the discussion has been in terms of models and abstract parameters. Now we introduce observations in more detail. We have already stressed that we in a given situation have a choice between different experiments/ questions $a$. In this section we will fix $a$, and hence fix the reduced parametric function $\lambda^a(\phi)$. Given a measurement instument, this will lead to a reduced model $\mathrm{P}^{\lambda^a}$. We will make some specific requirements - not too strong - on these models shortly. The sample space for all experiments will be called $S$, so that $\mathrm{P}^{\lambda^a}$ is a measure on $S$.

In this section we will need to introduce some statistical concepts; for a more thorough treatment, see, e.g., [18].

A random variables containing all the information of relevance to the particular experiment $a$, is called a sufficient random variable for this experiment, or a sufficient statistic. The concept of sufficiency has proved to be very useful in statistics. Precisely, we have the following

**Definition 3**
*A random variable $t^a = t^a(\omega); \omega \in S$ connected to a model $\mathrm{P}^{\lambda^a}$ is called sufficient if the conditional distribution of each other variable $y$, given $t^a$, is independent of the parameter $\lambda^a$.*

This means that all information about the parameter is contained in $t^a$. In general, $t^a$ will be a vector variable. A sufficient statistic (random variable) $t^a$ is minimal if all other sufficient statistics are functions of $t^a$. It is complete if

$$\mathrm{E}^{\lambda^a}(h(t^a)) = 0 \text{ for all } \lambda^a \text{ implies } h(t^a) \equiv 0. \tag{12}$$

It is well known that a minimal sufficient statistic always exists and is unique except for invertible transformations, and that every complete sufficient statistic is minimal. If the statistical model has a density belonging to an exponential class

$$b(y)d(\lambda)e^{c(\lambda)' t^a(y)},$$

and if $c(\Lambda) = \{c(\lambda) : \lambda \in \Lambda\}$ contains some open set, then the statistic $t^a$ is complete sufficient.

Recall from Definition 1 that a function $\xi(\lambda^a)$ is called unbiasedly estimable if $\mathrm{E}^{\lambda^a}(y) = \xi(\lambda^a)$ for some $y$. Given a complete sufficient statistic $t^a$, every unbiasedly estimable function $\xi(\lambda^a)$ has one and only one unbiased estimator that is a function of $t^a$. This is the unique unbiased estimator with minimum risk under weak conditions [18]. Thus complete sufficiency leads to efficient estimation.

**Definition 4.**
*Assume that a complete sufficient statistic $t^a$ exists under the model $\mathrm{P}^{\lambda^a}$. Let the model be dominated, i.e., such that all $\mathrm{P}^{\lambda^a}$ are absolutely continuous with respect to a common measure $\mathrm{P}$. Then the Hilbert space $\mathcal{K}^a$ is defined as consisting of all functions $h(t^a)$ such that $h(t^a) \in L^2(S, \mathrm{P})$.*

Let then $\dot{G}$ be the group acting upon the sample space $S$.

**Proposition 1.**
*Each space $\mathcal{K}^a$ is an invariant space for the regular representation of the observation group $\dot{G}$.*

Proof. If $t^a$ is sufficient under the model $\mathrm{P}^{\lambda^a}$, and $\dot{G}$ is the group on the sample space, then $t^a\dot{g}$ given by $t^a\dot{g}(\omega) = t^a(\omega\dot{g})$ is sufficient for all $\dot{g} \in \dot{G}$. This is proved by a simple excercise using (14) below. Also, if $t^a$ is complete, then $t^a\dot{g}$ must be complete; hence the two must be equivalent. Therefore $\mathcal{K}^a$ is invariant under $\dot{G}$.

Consider now the operator $A^a$ from $L^2(S, \mathrm{P})$ to $V_\lambda^a \subset L^2(\Phi, \nu))$ defined by

$$(A^a y)(\lambda^a(\phi)) = \int \mathrm{P}^{\lambda^a(\phi)}(d\omega) y(\omega) = \mathrm{E}^{\lambda^a(\phi)}(y), \tag{13}$$

using again the reduced model $\mathrm{P}^{\lambda^a}(d\omega)$ corresponding to the experiment $a$.

**Definition 5.**
*Define the space $\mathcal{H}^a$ by $\mathcal{H}^a = A^a\mathcal{K}^a$.*

By the definition of a complete sufficient statistic, the operator $A^a$ will have a trivial kernel as a mapping from $\mathcal{K}^a$ onto $A^a\mathcal{K}^a$. Hence this mapping is one-to-one. It is also continuous and has a continuous inverse. Hence $\mathcal{H}^a$ is a closed subspace of $\mathrm{L}^2(\Phi, \nu)$, and therefore a Hilbert space. Note also that $\mathcal{H}^a$ is the space of unbiasedly estimable functions with estimators in $\mathrm{L}^2(S, P)$. It is of course included in the space $V_\lambda^a$ of all functions of the parameter $\lambda^a$.

**Proposition 2.**
*The space $\mathcal{H}^a$ is an invariant space for the regular representation of the group $\bar{G}^a$.*

Proof.

Assume that $\xi(\lambda^a) = \mathrm{E}^{\lambda^a}(y)$ is unbiasedly estimable. Then also $\eta(\lambda^a) = \xi(\lambda^a \bar{g}) = \mathrm{E}^{\lambda^a \bar{g}}(y) = \mathrm{E}^{\lambda^a}(y\dot{g}^{-1})$ is unbiased estimable, so $\mathcal{H}^a$ is an invariant space under the regular representation $\bar{U}$ of $\bar{G}^a$.

**Theorem 1.**
*The two spaces $\mathcal{K}^a$ and $\mathcal{H}^a$ are unitarily related. Also, the regular representations of the groups $\dot{G}$ and $\bar{G}^a$ on these spaces are unitarily related.*

Proof.
We will show that the mapping $A^a$ can be replaced by a unitary map in the relation $\mathcal{H}^a = A^a\mathcal{K}^a$.

Recall that the connection $\dot{g} \mapsto \bar{g}$ from the observation group to the parameter group $\bar{G}^a$ is given from the reduced model by

$$\mathrm{P}^{\lambda^a \bar{g}}(B) = \mathrm{P}^{\lambda^a}(B\dot{g}^{-1}). \tag{14}$$

For $\dot{g} \in \dot{G}$ and $\bar{g} \in \bar{G}^a$ define $U_1(\dot{g}) = \bar{U}(\bar{g})$ as operators on $\mathcal{H}^a$ when $\dot{g} \mapsto \bar{g}$ as in (14). Here $\bar{U}$ is the regular representation of the group $\bar{G}^a$. Then it is easy to verify that $U_1$ is a representation of $\dot{G}$. Also, if $V_1$ is an invariant space for $U_1$, then it is also an invariant space for $\bar{U}$. However, the space $V_1$ is not necessarily irreducible for $\bar{U}$ even if it is irreducible for $U_1$.

Using the definition (13) and the connection (14) between $\dot{g}$ and $\bar{g}$ we find the following relationships. We assume that the random variable $y(\cdot)$ belongs to $\mathcal{K}^a \subset L^2(S, \mathrm{P})$ and that $\bar{U}$ is chosen as a representation on the invariant space $\mathcal{H}^a$. Then

$$\begin{aligned} U_1(\dot{g})A^a y(\lambda^a) = \bar{U}(\bar{g})A^a y(\lambda^a) = \textstyle\int y(\omega)\mathrm{P}^{\lambda^a \bar{g}}(d\omega) \\ = \textstyle\int y(\omega)\mathrm{P}^{\lambda^a}(d\omega\dot{g}^{-1}) = \int y(\omega\dot{g})\mathrm{P}^{\lambda^a}(d\omega) = A^a\dot{U}(\dot{g})y(\lambda^a), \end{aligned} \tag{15}$$

where $\dot{U}$ is the representation on $\mathcal{K}^a$ given by $\dot{U}y(\omega) = y(\omega\dot{g})$, i.e., the regular representation on $\mathrm{L}^2(S, \mathrm{P})$ restricted to this space.

Thus $U_1(\dot{g})A^a = A^a\dot{U}(\dot{g})$ on $\mathcal{K}^a$.

Furthermore

$$U(g) = \bar{U}(\bar{g}) = U_1(\dot{g}) = A^a\dot{U}(\dot{g})A^{a\ -1} \textit{ when } \dot{g} \mapsto \bar{g} \textit{ and } g \mapsto \bar{g}.$$

Recall that $g \mapsto \bar{g}$ in this setting if $(\lambda^a\bar{g})(\phi) = \lambda^a(\phi g)$, and that $\bar{U}(\bar{g}) = U(g)$ in this case. Furthermore, $U(g)f(\phi) = f(\phi g)$ when $f \in V_\lambda^a$ and $g \in G^a$.

By [38] p. 48, if two representations of a group are equivalent, they are unitary equivalent; hence for some unitary $C^a$ we have

$$\bar{U}(\bar{g}) = C^a\dot{U}(\dot{g})C^{a\ \dagger} \tag{16}$$

when $\dot{g} \mapsto \bar{g}$.

Since the unitary operators in this proof are defined on $\mathcal{K}^a$ and $\mathcal{H}^a$, respectively, it follows that these spaces are related by $\mathcal{H}^a = C^a\mathcal{K}^a$.

From a statistical point of view it is very satisfactory that the sufficient statistic determines the Hilbert space for single experiments. The sufficiency principle, by many considered to be one of the backbones of statistical inference (e.g. [39]) says that identical conclusions should be drawn from all sets of observations with the same sufficient statistic. It is also of importance that this Hilbert space satisfies the invariance properties that are needed in order that it can serve as a representation space for the symmetry groups connected to each experiment.

Definition 4 may also be coupled to the operator $A^a$ and to an arbitrary Hilbert space $\mathcal{K}'$ of sufficient statistics, which may trivially be the whole space $L^2(S, \mathrm{P})$. Let first

$$L^a = \{y \in \mathcal{K}' : \mathrm{E}^{\lambda^a} y = 0 \text{ for all } \lambda^a\}. \tag{17}$$

Then $\mathcal{K}^a$ may be considered as the factor space $\mathcal{K}'/L^a$, i.e., the equivalence classes of the old $\mathcal{K}'$ with respect to the linear subspace $L^a$ (cf [38], I.2.10IV).

Here is a proof of this fact: Let $\xi \in A^a\mathcal{K}'$, such that $\xi(\lambda^a) = \mathrm{E}^{\lambda^a}(y)$ for some $y \in \mathcal{K}'$. Then $y$ is an unbiased estimator of the function $\xi(\lambda^a)$. By [18], Lemma 1.10, $\xi(\lambda^a)$ has one and only one unbiased estimator which is a function $h(t^a)$ of $t^a$. Then every unbiased estimator of $\xi(\lambda^a)$ is of the form $y = h(t^a) + x$, where $x \in L^a$; this constitutes an equivalence class. On the other hand, every $h(t^a)$ can be taken as such a $y$.

# 14 The quantumtheoretical Hilbert space.

Our task in this section is to tie the spaces $\mathcal{H}^a$ together. We have already assumed that all the different potential experiments $a \in \mathcal{A}$ can be tied to one single observational space. This is a basic assumption in many macroscopic situations also; say the case where one must choose one of several potential treatments for one patient. In quantum mechanics one must assume an experimental setting such that a limitation of the complete sufficient statistics for experiment $a$ makes the observational Hilbert space $\mathcal{K}^a$ non-trivial. If these sufficient statistics should be related in some sense, this would mean intuitively that we have limited resources in the same way for the different experiments.

What we assume here, is that the reduced parameter spaces of the different experiments have a similar structure. Then the corresponding groups $G^a$ can be expected to be isomorphic. Precisely, we will assume an inner isomorphism as follows:

**Assumption 1.**

*Let $a, b$ be any pair of experiments. Assume then that there exists a group element $g_{ab} \in G$ such that the isomorphism between $G^a$ and $G^b$ is given by*

$$g^a = g_{ab} g^b g_{ab}^{-1}. \tag{18}$$

Here are some examples where this assumption is satisfied:

1) Let $\Phi$ be the real line, let $G$ be the reflection and translation group on $\Phi$, and let $g^a$ be the reflection around $a \in \Phi$, which together with the

identity constitutes the subgroup $G^a$. Then (18) holds if $g_{ab}$ is the translation $x \mapsto x + (b - a)$.

2) In the spin 1/2 case $\Phi$ was a space of vectors, $G$ was the rotation group, and $G^a$ was the subgroup of rotations around the axis $a$ together with a reflection around any axis perpendicular to $a$. Then (18) holds if $g_{ab}$ is any rotation transforming $a$ to $b$.

3) (See [10], p. 24) Let $\Phi$ be spacetime $\{\xi_1, \xi_2, \xi_3, \tau\}$. A Lorentz boost in the $\xi$ direction with velocity $v$ is given by the transformation (29) below. Call this transformation group element $g^v$, and let $g^{v,\eta,\sigma}$ be the corresponding boost taking the space time point $(\eta, \sigma)$ as an origin instead of $(0,0)$. Let $h^{\eta,\sigma}$ be a translation in spacetime by the amount $(\eta, \sigma)$. Then if $(\eta, \sigma) = (\xi, \tau) g^v$, we have the three relations

$$g^{v,\eta,\sigma} = h^{\eta,\sigma} g^v h^{\eta,\sigma\ -1},$$

$$h^{\eta,\sigma} = g^v h^{\xi,\tau} g^{v\ -1},$$
$$g^{u,\eta,\sigma} = g^v g^{u,\xi,\tau} g^{v\ -1}.$$

4) If (18) holds for transformations on some component spaces, it also holds for the cartesian product of these spaces when the relevant cartesian product of groups are used.

Assumptions 1 will be crucial in connecting the Hilbert spaces $\mathcal{H}^a$ for the different experiments.

From (18) follows that any representation $U$ of the basic group $G$ satisfies

$$U(g^a) = U(g_{ab}) U(g^b) U(g_{ab})^\dagger. \tag{19}$$

In particular, this is true for the regular representation $U$, which satisfies $U(g^a) = \bar{U}(\bar{g}^a)$ on $\mathcal{H}^a$ and $U(g^b) = \bar{U}(\bar{g}^b)$ on $\mathcal{H}^b$.

Since $\mathcal{H}^a$ and $\mathcal{H}^b$ are invariant spaces for these respective representations by Proposition 2, it follows that we can construct a connection between the spaces by

$$\mathcal{H}^a = U(g_{ab}) \mathcal{H}^b. \tag{20}$$

It follows from (20):

**Theorem 2.**

*a) There is a Hilbert space $\mathcal{H}$, and for each $a$ a unitary transformation $D^a$ such that $\mathcal{H}^a = D^a \mathcal{H}$.*

*b) There are unitary transformations $E^a$ such that the observational Hilbert spaces satisfy $\mathcal{K}^a = E^a \mathcal{H}$.*

Proof.

a) Obvious from (20).

b) From a) and Theorem 1.

**Theorem 3.**

*$\mathcal{H}$ is an invariant space for a representation of the whole group $G$.*

It follows from Proposition 2 that $\mathcal{H}^a$ is an invariant space for the group $\bar{G}^a$, hence for $G^a$. (Remember that $U(g^a) = \bar{U}(\bar{g}^a)$.)

This can now be extended. Assume that $g = g_1 g_2 g_3$, where $g_1 \in G^a$, $g_2 \in G^b$ and $g_3 \in G^c$. Then

$$D^{a\ \dagger} U^a(g_1) D^a D^{b\ \dagger} U^b(g_2) D^b D^{c\ \dagger} U^c(g_3) D^c$$

gives a representation on $\mathcal{H}$ of the set of elements in $G$ that can be written as a product $g_1 g_2 g_3$ with $g_1 \in G^a$, $g_2 \in G^b$ and $g_3 \in G^c$.

Continuing in this way, using the assumption that the group $G$ is generated by $\{G^a; a \in A\}$ we are able to construct a representation of the whole group $G$ on the space $\mathcal{H}$. In particular, one must be able to take $\mathcal{H}$ as an invariant space for a representation of this group.

As an example, the Hilbert space of a particle with spin is always an (irreducible) invariant space for the rotation group. Together with the fact that $\mathcal{H}$ should be an invariant space for the sample space group $\dot{G}$, this to a large extent determines $\mathcal{H}$, at least if the experimental setting forces $\mathcal{H}$ to be as small as possible.

# 15 Operators and states.

So, by what has just been proved, for each $a$ the Hilbert space $\mathcal{H}^a$ of unbiasedly estimable functions of $\lambda^a$ can be put in unitary corresponance with a common Hilbert space $\mathcal{H}$. In this Section we will make the assumption that the reduced parameter $\lambda^a$ only takes a finite or countable set of values. For simplicity we consider only the case where $\lambda^a$ is a scalar parameter.

Recall that $\mathcal{H}^a = D^a \mathcal{H}$ is a space of functions of the parameter $\lambda^a$. Define in general an operator $S^a$ on this space by

$$S^a f(\lambda^a) = \lambda^a f(\lambda^a) \tag{21}$$

for functions $f$ such that the function on the righthand side belongs to $\mathcal{H}^a$.

The corresponding operator on $\mathcal{H}$ will then be defined by

$$T^a = D^{a\dagger} S^a D^a. \tag{22}$$

Now $S^a$ has eigenvalues $\lambda_k^a$ with corresponding eigenfunction given by the rather trivial function $f_k^a$ which equals 1 when $\lambda^a = \lambda_k^a$, otherwise 0. This implies the important consequence that $T^a$ also has eigenvalues $\lambda_k^a$ with some eigenvector $v_k^a$. We will here and in the following for simplicity assume non-degenerate eigenvalues.

Note that $f_k^a$ is just the indicator for the statement $\lambda^a = \lambda_k^a$. Transferred to another space, this means that the eigenvectors $v_k^a$ of $\mathcal{H}$ also can be interpreted as an indicator of the same statement. Thus these vectors can be given the following interpretation: A question $a \in \mathcal{A}$ has been asked, and the answer is given by $\lambda^a = \lambda_k^a$. This is consistent with the well known quantum mechanical interpretation of a state vector.

To follow up, a natural conjecture of the present paper is that, since all vectors in $\mathcal{H}$ are eigenvectors of some operator, all pure 'states', expressed in quantum theory as such vectors, can be given an interpretation of this kind.

The operator $T^a$ may be written

$$T^a = \sum_{k=1}^{n} \lambda_k^a v_k^a v_k^{a\,\dagger}. \tag{23}$$

If $\lambda^a$ is multidimensional, a similar statement holds for multivariate operators and multidimensional eigenvalues.

**Theorem 4.**

*Under the assumptions above, the space $\mathcal{H}^a$ of unbiasedly estimable functions is equal to the space of estimable functions.*

Proof

If $T^a$ is an operator of the form (23), then $\xi(T^a) = \sum_{k=1}^n \xi(\lambda_k^a) v_k^a v_k^{a\,\dagger}$ is also a valid operator for any function $\xi$. Thus $\xi(S^a)$, having eigenvalues $\xi(\lambda_k^a)$ also has eigenvectors in $\mathcal{H}^a$ which are functions of $\lambda^a$. This means that $\xi(\lambda^a)$ is unbiasedly estimable for any $\xi$.

More on the relationship between the foundation of quantum mechanics on the one hand and the group representations and their operators on the other hand, can be found in [10].

An important question, probably requiring deeper mathematical tools than what has been used here, is to generalize the results of this paper to Hilbert spaces of infinite dimension.

# 16 The Born formula.

To complete deriving the formalism of quantum mechanics from the statistical parameter approach the most important task left is to arrive at the Born formula

$$\mathrm{P}(\lambda^b = \lambda_j^b | \lambda^a = \lambda_k^a) = |v_k^{a\dagger} v_j^b|^2. \tag{24}$$

Note that here $\lambda^a$ and $\lambda^b$ are connected to two different experiments on the same unit(s). The interpretation of (24) is as follows: Assume that the system is in the state where the reduced parameter corresponding to $\lambda^a$ is equal to $\lambda_k^a$. Then imagine that we will perform an experiment whose outcome $y$ depends on the the parameter $\lambda^b$. The limited experimental basis again enforces a reduced model, and in an ideal experiment, where observation almost equals parameter, the probability distribution according to quantum mechanics is given by (24), where $v_k^a$ and $v_j^b$ are the corresponding eigenvectors. To simplify the discussion, we have assumed non-degenerate eigenvalues.

The proof of (24) may be related to that of Gleason's theorem [40]; for a formulation see also for instance [41]. The derivation of a probability law from Gleason's theorem has been argued for by [42], but note that this theorem takes a different set of assumptions as a point of departure.

A more direct approach to (24) using decision theory has been given by Deutsch [43]. Deutsch started by making reasonable assumptions about how a rational decision maker should behave, and then proceded from simple games to more complicated situations. Several of his arguments were heuristic. On this background the paper has been criticized by Gill [44] who added three assumptions: Degeneracy in eigenstates, functional invariance and unitary invariance. He conjectured that Born's formula can be proved under these assumptions. The assumptions of Gill are satisfied by the theory of the present paper. For instance the theory, including the definition (23) is invariant under the transformations $T^a \mapsto U^\dagger T^a U$ and $\mathcal{H} \mapsto U\mathcal{H}$ for some fixed unitary $U$.

The paper by Deutch [43] has also been criticized by Finkelstein [45]. Further developments are given in [42, 46, 47]. The final solution is still wanting.

## 17 Basis of quantum mechanics and link to quantum statistics.

Our state concept may now be summarized as follows: To the state $\lambda^a(\cdot) = \lambda_k^a$ there corresponds the state vector $v_k^a$, and these vectors determine the transition probabilities as in (24). The probability distribution (24) also implies

$$\mathrm{E}(\lambda^b|\lambda^a = \lambda_k^a) = v_k^{a\dagger} T^b v_k^a, \tag{25}$$

where $T^b = \sum \lambda_j^b v_j^b v_j^{b\dagger}$. Similarly

$$\mathrm{E}(f(\lambda^b)|\lambda^a = \lambda_k^a) = v_k^{a\dagger} f(T^b) v_k^a, \tag{26}$$

where $f(T^b) = \sum f(\lambda_j^b) v_j^b v_j^{b\dagger}$. Thus, in ordinary quantum mechanical terms, the expectation of every observable in any state is given by the familiar formula.

It follows from (25) and from the preceding discussion that the first three rules of [41], p. 71, taken there as a basis for quantum mechanics, are satisfied. The 4th rule, the Schrödinger equation, will be discussed below. The present approach may also be related to the seven principles of quantum mechanics put forward by Volovich [48], but this will require some further developments.

In ordinary statistics, a measurement is a probability measure $\mathrm{P}^\theta(dy)$ depending upon a parameter $\theta$. Assume now that such a measurement depends upon the parameter $\lambda^b(\cdot)$, while the current state is given by $\lambda^a(\cdot) = \lambda_k^a$. Then as in (26), for each element $dy$ there exist an operator $M(dy)$ such that

$$\mathrm{P}[dy|\lambda^a = \lambda_k^a] = v_k^{a\dagger} M(dy) v_k^a, \text{ namely } M(dy) = \sum_j \mathrm{P}^{\lambda_j^b}(dy) v_j^b v_j^{b\dagger}.$$

As is easily checked, these operators satisfy $M[S] = I$ for the whole sample space $S$, and furthermore $\sum M(A_i) = M(A)$ for any finite or countable sequence of disjoint elements $\{A_1, A_2, \ldots\}$ with $A = \cup_i A_i$.

A more general state assumption is a Bayesian one corresponding to this setting: Let the current state be given by probabilities $\pi(\lambda_k^a)$ for different values of $\lambda_k^a$. Then, defining $\sigma = \sum \pi(\lambda_k^a) v_k^a v_k^{a\dagger}$, we get

$$\mathrm{P}[dy] = \mathrm{tr}[\sigma M(dy)].$$

This is the basis for much of quantum theory, in particular for the quantum statistical inference in [49]; for a formulation, see also [41].

## 18 Isolated systems.

We now have a rather complete description of a large class of isolated systems. These systems can be in a state described by a state vector $v_k^a$ in a Hilbert space $\mathcal{H}$, which means that an ideal experiment $a \in \mathcal{A}$ has been performed with the result expressed in terms of the (reduced) parameter as $\lambda^a = \lambda_k^a$. An alternative interpretation: *If* experiment $a$ should be performed, the result will be $\lambda^a = \lambda^a_k$ with certainty.

We assume that every state of a completely isolated system is such an eigenstate, meaning that it is equivalent to some statement $\lambda^a = \lambda^a_k$.

This may be taken as consistent with the following empirical fact: For real quantum mechanical systems, all states are eigenstates for variables that are absolutely conserved, i.e. charge and mass. Linear combinations of such state vectors do not correspond to anything in reality, the well known phenomenon of superselection rules.

Also, when a system has an absolute symmetry $g$, i.e., identical particles under permutation, then the state vector has the corresponding symmetry, that is $U(g)v^a_k = v^a_k$. This is related to the Bose-Einstein statistics and the Fermi-Dirac statistics.

The state vector is only defined modulo a phase factor. This can be related to a non-trivial stability group for the group $G$ governing the system. In particular, this implies the following: Assume that there are parameters $\xi$ that are constant under the actions of the group $G$. Then each state vector $v^a_k$ describes the same state as any vector of the form $\exp[iF(\xi)]v^a_k$.

Constructing the joint state vector for a system consisting of several partial systems, with symmetries only within the partial systems, follows the receipt $v^{a_1a_2a_3}_{i_1i_2i_3} = v^{a_1}_{i_1} \otimes v^{a_2}_{i_2} \otimes v^{a_3}_{i_3}$.

# 19 The Lorentz transformation and Planck's constant.

We continue to insist upon keeping the distinction between ideal values of variables, that is, parameters on the one hand, and observed values on the other hand. In the statistical traditions we will continue to denote the former by greek letters. Hence let $(\xi_1, \xi_2, \xi_3)$ be the ideal coordinates of a particle at time $\tau$, and let $(\pi_1, \pi_2, \pi_3)$ be the (ideal) momentum vector and $\epsilon$ the (ideal) energy. In this section we will not speak explicitly about observations. Nevertheless it is important to be reminded of the premise that these quantities are theoretical, and that each single of them can only be given a concrete value through some given observational scheme.

This is a general way of thinking which also seemingly may serve to clarify some of the paradoxes of quantum theory. As an example, look at the Einstein, Podolsky, Rosen [1] situation in its original form: Two particles have position $\xi^i$ and momentum $\pi^i$ $(i = 1, 2)$. Since the corresponding quantum operators commute, it is in principle possible to have a state where both $\xi^1 - \xi^2$ and $\pi^1 + \pi^2$ are accurately determined. That implies that a measurement of $\xi^1$, respectively $\pi^1$ at the same time gives us accurate information on $\xi^2$, respectively $\pi^2$. We have a free choice of which measurement to make at the 1-particle, but that does not mean that this choice in any way influences anything at the 2-particle. It only influences which information we extract about this particle.

After this digression we continue with the single particle situation. As is well known from special relativity, the four-vectors $\xi = (\xi_1, \xi_2, \xi_3, \xi_0 = c\tau)$ and $\pi = (\pi_1, \pi_2, \pi_3, \pi_0 = c^{-1}\epsilon)$ transform according to the extended Lorentz transformation, the Poincaré transformation, which is the group which fixes $c^2 d\tau_0^2 = c^2 d\tau^2 - \sum_{i=1}^{3} d\xi_i^2$, respectively $c^2 m_0^2 = c^{-2}\epsilon^2 - \sum_{i=1}^{3} \pi_i^2$. This is a group of static linear orthogonal transformations of vectors together with the transformation between coordinate frames having a velocity $v$ with respect to each other. Specifically, the coordinate vectors transform according to an inhomogeneous transformation $\xi \mapsto A\xi + b$, while the momentum vector transforms according to the corresponding homogeneous transformation $\pi \mapsto A\pi$. The group might be a natural transformation group to link to the eightdimensional parameter $\phi = (\xi_1, \xi_2, \xi_3, \tau, \pi_1, \pi_2, \pi_3, \epsilon)$, associated with a particle at some time $\tau$. However, since the static rotations have representations associated with angular momenta already briefly discussed, we limit ourselves here to the group $G$ of translations together with the pure Lorentz group.

Consider then the groups $B_j$ given for $g_j^b \in B_j$ by $\xi_j g_j^b = \xi_j + b$, other coordinates constant, and the groups $V_j$ given by Lorentz boosts of some size $v$ in the direction of the coordinate axis of $\xi_j$ for $j = 1, 2, 3$ together with the time translation group $B_0$ given by $\tau g_0^t = \tau + t$. These groups generate $G$, and they are all abelian. Furthermore, the groups $B_j$ commute among themselves, the groups $V_j$ commute among themselves, and, since lengths perpendicular to the direction of the Lorentz boost are conserved, $B_j$ commute with $V_k$ when $j \neq k$. Finally, the elements of the group $B_0$ commutes with those of $B_j$ $(j \geq 1)$, but not with those of $V_j$ $(j = 1, 2, 3)$.

Disregarding the time translation group for a moment, it is left to consider, say, the groups $B_1$ and $V_1$ together. As is easily seen from the formula, these do not commute. The simplest one is $B_1$, which only affects the coordinate $\xi_1$. Hence $\xi_1$ is trivially permissible with respect to this group.

From the form of the Lorentz transformation

$$\xi_1 \mapsto \frac{\xi_1 + v\tau}{\sqrt{1 - (\frac{v}{c})^2}}, \quad \tau \mapsto \frac{\tau + \frac{v}{c^2}\xi_1}{\sqrt{1 - (\frac{v}{c})^2}} \tag{27}$$

and correspondingly for $(\pi_1, \epsilon)$, we see that $\xi_1$ and $\pi_1$ are not permissible when $\tau$, respectively $\epsilon$ are variable. The linear combinations $\xi_1 - c\tau$, $\xi_1 + c\tau$, $\pi_1 - c^{-1}\epsilon$ and $\pi_1 + c^{-1}\epsilon$ are permissible. One could conjecture that these facts could be useful in a relativistic quantum mechanics, but this will not be pursued here.

Furthermore, $V^\xi = \{f : f(\phi) = q(\xi_1(\phi)) \text{ for some } q\}$ is a subspace of $L^2(\Phi, \nu)$ which is invariant under the group $G_1^b$. The representations have the form $U_1(g)q(\xi_1) = q(\xi_1 g) = q(\xi_1 + b)$. But

$$q(\xi_1 + b) = \sum_{k=0}^{\infty} \frac{b^k}{k!} \frac{\partial^k}{\partial \xi_1^k} q(\xi_1) = \exp(b\frac{\partial}{\partial \xi_1}) q(\xi_1) = \exp(\frac{ibP_1}{\hbar}) q(\xi_1),$$

where $P_1$ is the familiar momentum operator

$$P_1 = \frac{\hbar}{i} \frac{\partial}{\partial \xi_1}$$

.

Thus the particular group formulated above has a Lie group representation on an invariant space with a generator equal to the corresponding momentum operator of quantum mechanics. The proportionality constant $\hbar$ can be argued to be the same for all momentum components (and energy) by the conservation of the 4-vector. By similarly considering systems of particles one can argue that $\hbar$ is a universal constant.

In particular then, time translation $\tau \mapsto \tau + t$ has a representation

$$\exp(\frac{iHt}{\hbar}), \tag{28}$$

where $H$ is the Hamiltonian operator.

All these operators can be connected to representations the group $G$ as defined above. In [10] is pointed out that the Lorentz transformation (27) is equivalent to

$$\xi_1 \mapsto \xi_1 \mathrm{cosh} r_v + c\tau \mathrm{sinh} r_v, \quad c\tau \mapsto \xi_1 \mathrm{sinh} r_v + c\tau \mathrm{cosh} r_v, \tag{29}$$

where the rapidity $r_v$ is defined by $\mathrm{tanh} r_v = v/c$. This makes the Lorentz boost additive in the rapidity, and all relevant operators and their commutation relations can be derived. In particular, the familiar commutation relation $X_1 P_1 - P_1 X_1 = i\hbar I$ (with $X_1$ being the operator corresponding to position $\xi_1$) holds under the approximation $r_v \approx v/c$.

The corresponding commutation relation between the time operator and the energy operator has also been derived by Tjøstheim [50] in a stochastic process setting using just classical concepts.

Starting from these commutation relations, other representations of this Heisenberg-Weil group are discussed in [51].

Note that the groups $G_1^b$ and $G_1^v$ are transitive in this case, so there is no need for - or possibility of - a model reduction.

In the nonrelativistic approximation, $\xi_1$ and $\pi_1$ are permissible. The basis vectors of the Hilbert space for position $\xi_1$ and basis vectors of the Hilbert space for momentum $\pi_1$ are connected by a unitary transformation of the form

$$u^\pi(\pi_1) = \frac{1}{\sqrt{2\pi\hbar}} \int \exp(\frac{i\pi_1\xi_1}{\hbar}) u^\xi(\xi_1) d\xi_1.$$

The parameters $\xi_1$ and $\pi_1$ can be estimated by making observations. It is natural to impose the translation/ Lorentz group upon these measurements. Thus the requirement that the basic Hilbert space also should be a representation space for the observation group, is obviously satisfied in this case.

# 20 Time development. Schrödinger equation.

In Section 19 we showed that in the case of a single particle, the time translation $\tau \mapsto \tau + t$ had the group representation

$$\exp(\frac{iHt}{\hbar}), \tag{30}$$

where $H$ is the Hamiltonian operator. This can be generalized to systems of several particles using an assumption of additive Hamiltonian, and assuming that the particles at some point of time were pairwise in contact, or at least so close with respect to space and velocity that relativistic time scale differences can be neglected. The operator (30) acts on the Hilbert space $\mathcal{H}$.

Assume further that at time 0 a maximal measurement is done, so that the system is in some state $v_0 \in \mathcal{H}$. This means, according to our interpretation, that some experiment with reduced parameter $\lambda^a$ has been done, resulting in a value $\lambda_1^a$. The construction of the Hilbert space $\mathcal{H}$ was carried out in the Sections 13-14; the starting point was $\mathcal{K}^a$, then $\mathcal{H}^a = A^a\mathcal{K}^a$, where $A^a$ was given by

$$A^a y(\lambda^a) = \mathrm{E}^{\lambda^a}(y),$$

but where one also has $\mathcal{H}^a = C^a\mathcal{K}^a$ for some unitary operator $C^a$. Note that from Theorem 2, $\mathcal{H}^a = D^a\mathcal{H}$ and $\mathcal{K}^a = E^a\mathcal{H}$ for unitary operators $D^a$ and $E^a$.

The vector $v_0$ corresponds to some vector $w_0$ in $\mathcal{K}^a$ by this unitary transformation, then to $u_0 = A^a w_0 \in \mathcal{H}^a$.

Consider now the time translation group element with step $t$, and assume that $\lambda^a$ transforms under this group element into a new parameter $\lambda^a(t)$. By the regular representation of the time translation group, this leads to a new operator $A^{a,t}$ given by

$$A^{a,t}y(\lambda^a) = A^a y(\lambda^a(t)) = \exp(\frac{iH_1 t}{\hbar})A^a y(\lambda^a). \tag{31}$$

Here $H_1 = D^a H D^{a\ -1}$ is the Hamilton operator $H$ transformed from the basic Hilbert space $\mathcal{H}$ to the parameter space $\mathcal{H}^a = D^a\mathcal{H}$ for experiment $a$.

Vectors $u(t)$ in the space $A^{a,t}\mathcal{K}^a$ correspond to vectors $(A^{a,t})^{-1}u(t)$ in $\mathcal{K}^a$. In particular, then, during the time span $t$, we have that $w_0$ in $\mathcal{K}^a$ develops into

$$w_t = (A^a)^{-1}\exp(-\frac{iH_1 t}{\hbar})A^a w_0.$$

Since the operator $\exp(-iH_1 t/\hbar)$ is unitary, it follows again from [38], p. 48 that $A^a$ can be replaced by a unitary operator:

$$w_t = (C^a)^{-1}\exp(-\frac{iH_1 t}{\hbar})C^a w_0 = \exp(-\frac{i(C^a)^{-1}H_1 C^a t}{\hbar})w_0.$$

Transforming back from $\mathcal{K}^a$ to $\mathcal{H}$, the state vector at time $t$ will be

$$v_t = \exp(-\frac{iHt}{\hbar})v_0.$$

As is well known, the latter equation is just a formulation of the familiar Schrödinger equation

$$i\hbar\frac{\partial}{\partial t}v_t = Hv_t.$$

# 21 Paradoxes and some further themes.

Here I include a very brief discussion of some familiar themes from quantum mechanics, many of which are discussed in several textbooks. A very recent discussion of several points is given in Laloë [52].

Of course, much more can be said on each theme. Some of the statements below are controversial, and many are certainly too simplified. The brief statements may serve as a starting point of a discussion, however. Our main concern is to point out similarities between (our version of) quantum physics and statistical modelling.

The status of the state vector.

We concentrate on a discrete parameter, typically multidimensional: Suppose that $\lambda^a(\phi)$ is maximal in the sense that no parameter can be connected to any experiment in such a way that $\lambda^a$ is a function of this parameter. Then the operator $T^a$ corresponding to $\lambda^a(\cdot)$ has a non-degenerate spectrum. Thus each specification $(\lambda^a(\phi) = \lambda^a_k)$ is equivalent to specifying a single vector $v^a_k$. We emphasize that $\lambda^a$ is a parameter which is specifically connected to the experiment (or question) $a \in \mathcal{A}$.

Thus in this case the state can be specified in two equivalent ways. For a non-physicist the specification $(\lambda^a(\phi) = \lambda^a_k)$, that is, specifying all quantum numbers, is definitively simpler to understand than the Hilbert vector specification. It is easy to see that every Hilbert space vector is the eigenvector of *some* operator. Assuming that this operator can be chosen to correspond to some $\lambda^a$, it then follows that the state vector can be written as equivalent to some $(\lambda^a(\phi) = \lambda^a_k)$. A more general statement will include continuous parameters.

A limitation of this, however, is for a state evolving through the Schrödinger equation. While it might be true that $v_t$ at each $t$ is equivalent to *some* statement $(\lambda^a(t) = \lambda^a_k(t))$, the parameter $\lambda^a(t)$ will in case necessarily change with time.

Note also, of course, that in the formulae of Section 17 and in related results, the state vector is needed explicitly.

Collapse of the wave packet.

If we maintain that the rôle of the wavefunction is to give condensed information about what is known about one or several parameters of the system, then it is not strange that the wavefunction changes at the moment when new such information is obtained. Such a collapse due to change of information is well known in statistics.

Superselection rules.

When parameters are absolutely conserved, for instance charge or mass, then also in conventional quantum mechanics no linear combination is allowed between the vectors specifying the different corresponding states. This may to some extent serve to emphasize our view that a wave function makes sense only if it can be made equivalent to some statement $\lambda^a(\phi) = \lambda^a_k$.

Wigner's friend etc.

In principle a statistical model can be formulated for a given system either excluding a certain observer (measuring apparatus) in the model, or including this observer. There is no contradiction between these two points of view in principle.

Bohr complemetarity.

A limited experimental basis implies that an experimentalist must choose between measuring/ specifying the maximal parameter $\lambda^1$ or the maximal parameter $\lambda^2$. It is impossible to specify both. And knowledge of both parameters is impossible to have. As has been stated earlier, several macroscopic examples of the same phenomenon can be found.

Schrödinger's cat.

The total parameter $\phi$ can again be imagined to give a complete description of the whole system, including the death status of the cat. What can be observed in practice, is one of several complementary parameters $\lambda^a$, many of which include information of the death status, but some which don't. Included among the latter is the state variable developed by registrating the initial state of the radioactive source, and then letting some time go.

Decoherence.

When a system in a state $\lambda^a(\phi) = \lambda^a_k$ enters into an interaction with an environment with a large degree of freedom, a state involving a probability distribution over different $\lambda^a_k$-values will soon emerge.

Histories.

Choosing different focus parameters or experiments at each of a sequence of time points, we get a history of the kind

$$\lambda^1(\phi) = \lambda^1_i,\ \lambda^2(\phi) = \lambda^2_j,\ \lambda^3(\phi) = \lambda^3_k, \ldots .$$

The resulting sequence of quantum mechanical states has been discussed by Griffiths [53] and others.

Many worlds (Everett [5]).

There is a need for many models, not many wolds. Each time a choice of a measurement is made, a new model is needed. All future sequences of potential choices will make a need for many new models.

Quantum mechanics and relativity.

Relativistic quantum mechanics is beyond the scope of the present paper. However, it is well known that the use of symmetries, in particular representation theory for groups is much used in relativistic quantum mechanics and in elementary particle physics. Hence a development of the theory in that direction may appear to be possible, and would certainly be of interest.

It has often been said that it is difficult to reconcile quantum mechanics with general relativity theory. While this at the moment is mere speculations, one possible explanation may be that the transformation groups in general relativity are so large that no representation theory exists. (Say, the groups are not locally compact.) In that case the formal apparatus of quantum mechanics has no place. However, in principle it might still be that the present approach base on models, symmetry, focus parameters and model reduction may prove to be useful.

## 22 Concluding remarks.

The two most important arguments for the approach of the present paper, are as follows:

(1) Instead of taking formal, abstract axioms as the point of departure, we develop the theory using at each point comparatively reasonable assumptions.

(2) In principle the theory is an extension of current statistical theory under symmetry assumptions. Hence the concepts involved can be related to concepts that have proved useful also in other areas of science.

Having said this, it must also be said that there absolutly will be aspects of the theory as formulated in this paper, that will need to be developed further.

Note that the framework of the theory discussed here is very general. The total parameter $\phi$ can be almost anything, and contain any set of different parameters $\theta^a$. These aspects of the theory may also be relevant to large-scale statistics.

### Acknowledgements

I am grateful to Richard Gill for discussions and for showing genuine interest in the developments now shown of this paper. Also, comments on an earlier version by Peter McCullagh and encouragement from David Cox are appreciated.

The paper was inspired by two conferences in Växjö, arranged by Andrei Khrennikov and supported by the EU-network Quantum Probability with Applications to Physics, Information Theory and Biology, contract HPRN-CT-2002-00279.

## References

[1] Einstein, A., B. Podolski and N. Rosen, Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **47**, 777-780 (1935).

[2] Gödel, K., On Formally Undecidable Propositions of Principia Mathematica and Related Systems. Dover (1962).

[3] Bohr, N., Can quantum-mechanical description of physical reality be considered complete? Phys. Rev. **48**, 696-702 (1935).

[4] Bohm, D., A suggested interpretation of the quantum theory in terms of "hidden" variables. I. Phys. Rev. **85**, 166-179 (1952).

[5] Everett III, H., "Relative state" formulation of quantum mechanics. Rev. Mod. Phys. **29**, 454-462 (1957).

[6] Cramer, J.C., The transactional interpretation of quantum mechanics. Rev. Mod. Phys. **58** (1986), 647-687.

[7] Mermin, N.D., What is quantum mechanics trying to tell us? Am. J. Phys. **66**, 753-767 (1998).

[8] Fuchs, C.A., Quantum mechanics as quantum information. http://xxx.lanl.gov/abs/quant-ph/0205039 (2002).

[9] de Muynck, W.M., Towards a neo-Copenhagen interpretation of quantum mechanics. http://xxx.lanl.gov/abs/quant-ph/0307235 (2003).

[10] Bohr, A. and O. Ulfbeck, Primary manifistation of symmetry. Origin of quantal inderminacy. Rev. Mod. Phys. **67**, 1-35 (1995).

[11] Caves; C.M., C.A. Fuchs and R. Schack, Making good sense of quantum probabilities. http://xxx.lanl.gov/abs/quant-ph/0106133 (2001).

[12] Hardy, L., Quantum theory from five reasonable axioms. http://xxx.lanl.gov/abs/quant-ph/0101012 (2001).

[13] Helland, I.S., Quantum mechanics from symmetry and statistical modelling. Int. J. Theor. Phys. **38**, 1851-1882. (1999).

[14] von Neumann, J., Mathematische Grundlagen der Quantenmechanik. Julius-Springer-Verlag, Berlin (1932).

[15] Bell, J.S., On the problem of hidden variables in quantum mechanics. Rev. Mod. Physics **38**, 447-452 (1966).

[16] Kass, R.E. and L. Wasserman, The selection of prior distributions by formal rules. J. Amer. Stat. Ass. 91, 1343-1370 (1996).

[17] Petersen, A., 'The philosophy of Niels Bohr', in A.P. French and P.I. Kennedy [Eds.], Niels Bohr, A Centary Volume Harvard University Press, Cambridge, MA (1985).

[18] Lehmann, E.L. and G. Casella, Theory of Point Estimation. Springer, New York (1998).

[19] Helland, I.S., Statistical inference under a fixed symmetry group. Preprint, http://www.math.uio.no/ ingeh/publ.html (2002).

[20] McCullagh, P., What is a statistical model? Ann.Statistics **30**, 1225-1310 (2002).

[21] Helland, I.S., Discussion of P. McCullagh: What is a statistical model? Ann. Statistics **30**, 1225-1310 (2002).

[22] Helland, I.S., Extended statistical modelling under symmetry; the link towards quantum mechanics. Preprint. Available on http://folk.uio.no/ingeh/publ.html (2003).

[23] Helland, I.S., Reduction of regression models under symmetry. Contemporary Mathematics **287**, 139-153 (2001).

[24] Bell, J.S., Bertlmann's socks and the nature of reality. In: Speakable and unspeakable in quantum mechanics. Cambridge Univ. Press, Cambridge (1987).

[25] Accardi L., The quantum probabilistic approach to the foundation of quantum theory: Urns and chameleons. M.L. Dalla Chiara et al. [Eds.], Language, Quantum, Music, 95-104. Kluwer Academic Publishers (1999).

[26] Accardi,L. and M. Regoli, The EPR correlation and the chameleon effect. http://xxx.lanl.gov/abs/quant-ph/0110086 (2001).

[27] Gill, R., Accardi contra Bell (cum mundi): The impossible coupling.In: M. Moore, S. Froda and C. Láger (eds.) *Mathematical Statistics and Applications. Festschrift for Constance von Eeden*; IMS Lecture Notes - Monograph **42**, 133-154. Institute of Mathematical Statistics, Hayward, Ca. (2002).

[28] Gill, R.D., Time, finite statistics, and Bell's fifth position.In: *Foundations of Probability and Physics - 2*, Vol. 5 of Math. Modelling, Phys., Engin. and Cogn. Sc., Växjö University Press. Also: http://xxx.lanl.gov/abs/quant-ph/0301059 (2003).

[29] Thompson, C.H. and H. Holstein, The 'Chaotic Ball' model: local realism and the Bell test 'detection loophole'. http://xxx.lanl.gov/abs/quant-ph/0210150 (2002).

[30] Berger, J.O. and R.L. Wolpert, The Likelihood Principle. Institute of Mathematical Statistics. Hayward, California (1984).

[31] Cox, D.R., Some problems connected with statistical inference. Ann. Math. Statist. **29**, 357-372.

[32] Accardi, L., K. Imafuku and M. Regoli, On the physical meaning of the EPR-chameleon experiment. http://xxx.lanl.gov/abs/quant-ph/0112067 (2001).

[33] Gill, R.D., The chaotic chameleon http://www.math.uu.nl/people/gill (2003)

[34] Hamermesh, M., Group Theory and its Application to Physical Problems. Addison-Wesley, Reading, Massachusetts (1962).

[35] Wolbarst, A.B., Symmetry and Quantum Systems. Van Nostrand, New York (1977).

[36] Serre, J.-P., Linear Representations of Finite Groups. Springer-Verlag, Berlin (1977).

[37] Barut, A.S. and R. Raczka, Theory of Group Representation and Applications. Polish Scientific Publishers, Warsa, (1985).

[38] Naimark, M.A. and A.I. Štern, Theory of Group Representations. Springer-Verlag (1982).

[39] Lindsey, J.K., Parametric Statistical Inference. Clarendon Press, Oxford (1996).

[40] Gleason, A., Measures on closed subspaces of a Hilbert space. Journal of Mathematics and Mechanics **6**, 885-893 (1957).

[41] Isham, C.J., Lectures on Quantum Theory, Imperial College Press, London (1985).

[42] Barnum, H., C.M. Caves, J. Finkelstein, C.A. Fuchs and R. Schack. Quantum probability from decision theory? Proc. Roy. Soc. Land. Ser. A **456**, 1175-1182 (2000).

[43] Deutsch, D., Quantum theory of probability and decisions. Proc. Roy. Soc. **A 455**, 3129-3197 (1999).

[44] Gill, R., On an argument of David Deutsch. http://xxx.lanl.gov/abs/quant-ph/0307188 (2003).

[45] Finkelstein, J., Quantum probability from decision theory? http://xxx.lanl.gov/abs/quant-ph/9907004 (1999).

[46] Saunders, S., Derivation of the Born rule from operational assumptions. http://xxx.lanl.gov/abs/quant-ph/0211138 (2002).

[47] Wallace, D., Quantum probability and decision theory, revisited. http://xxx.lanl.gov/abs/quant-ph/0211104 (2002).

[48] Volovich, I.V., Seven principles of quantum mechanics. http://xxx.lanl.gov/abs/quant-ph/0212126 (2002).

[49] Barndorff-Nielsen, R., R. Gill and P.E. Jupp, On quantum statistical inference. J. Royal Statist. Soc. B **65**, 775-816. (2003)

[50] Tjøstheim, D., A commutation relation for wide sense stationary processes. SIAM. J. APPL. MATH. **30**, 115-122 (1976).

[51] Perelomov, A., Generalized Coherent States and Their Applications. Springer-Verlag, Berlin (1986).

[52] Laloë, F., Do we really understand quantum mechanics? Strange correlations, paradoxes and theorems. http://xxx.lanl.gov/abs/quant-ph/0209123 (2002).

[53] Griffiths, R.B., Consistent histories and the interpretation of quantum mechanics. Journal of Statistical Physics **36**, 219-274 (1984).